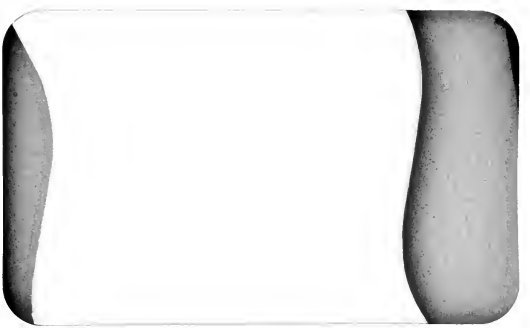
NUMERICAL METHODS FOR SOLVING INVERSE EIGENVALUE PROBLEMS*

Jorge Nocedal and Michael L. Overton

April 1983

Report No. 74



NUMERICAL METHODS FOR SOLVING INVERSE EIGENVALUE PROBLEMS*

Jorge Nocedal and Michael L. Overton

April 1983

Report No. 74

*This paper will appear in the Proceedings of the Third Inter-American Workshop on Numerical Methods (Caracas, Venezuela, June 1982), V. Pereyra, ed., Lecture Notes in Mathematics, Springer-Verlag.

NUMERICAL METHODS FOR SOLVING INVERSE EIGENVALUE PROBLEMS†

Jorge Nocedal and Michael L. Overton
Courant Institute of Mathematical Sciences
New York University
New York, N.Y. 10012

1. Introduction

Consider the following inverse eigenvalue problem:

Given n real symmetric n×n matrices $A_1,\ldots,A_n$ and n eigenvalues $\lambda_1^* \leq \ldots \leq \lambda_n^*$, find the coefficients $c_1,\ldots c_n$ such that the matrix

$$A(c) \triangleq c_1A_1+\ldots+c_nA_n \tag{1.1}$$

has the given eigenvalues. Here $c = [c_1,\ldots,c_n]^T$.

A slight reformulation of (1.1) describes the additive inverse eigenvalue problem, which arises in the solution of inverse Sturm-Liouville problems. In practice it happens frequently that only some eigenvalues are given. However, for the purpose of analysis it is convenient to consider the case where the number of parameters, eigenvalues and the order of the matrices is the same.

Another area where problem (1.1) arises is in shell model computations in nuclear spectroscopy (see Brussard and Glaudemans (1977)). There A is the Hamiltonian, the variables $\{c_i\}$ are the interactions of one and two bodies, and the matrices $\{A_i\}$ represent the result of adding and symmetrizing (or antisymmetrizing) the effect of many particles.

We will now briefly describe an inverse Sturm-Liouville problem. Consider the boundary value problem

$$-u'' + p(x)u = \lambda u \qquad \text{on } [0,\pi] \tag{1.2}$$

with some boundary conditions, for example $u(0) = u(\pi) = 0$. Suppose that p(x) is unknown. Given a spectrum $\{\lambda_i^*\}_1^\infty$ for the problem (1.2), can we determine p(x)? Let us discretize (1.2): let $h = \frac{\pi}{n+1}$, $u_i = u(ih)$, $p_i = p(ih)$, $i = 1,\ldots,n$.

†This work was supported in part by the Department of Energy, grant DEA CO276 ERO 3077-V, and in part by the National Science Foundation, grant MCS-81-01924.

Then, using finite differences to approximate u",

$$\frac{-u_{i+1} + 2u_i - u_{i-1}}{h^2} + p_i u_i = \lambda_i u_i \qquad (i=1,\ldots,n) \qquad (1.3)$$

$$u_0 = u_{n+1} = 0 .$$

In matrix form (1.3) can be written as

$$(A_0 + D)u = \lambda u \qquad (1.4)$$

where

$$A_0 = \frac{1}{h^2}\begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & & \\ & & & \ddots & \\ & & & & 2 \end{bmatrix}, D = \operatorname{diag}(p_i) , \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \quad \lambda = \begin{bmatrix} \lambda_1^* \\ \vdots \\ \lambda_n^* \end{bmatrix}$$

The problem is therefore: given A_0 and a spectrum $\{\lambda_i^*\}_1^n$ find a diagonal matrix D such that A_0+D has the given spectrum. This is the additive inverse eigenvalue problem. It can be written in the form

$$A(c) = A_0 + \sum_{i=1}^{n} c_i A_i \qquad (1.5)$$

with $A_i = e_i e_i^T$, where e_i is the i-th column of the identity matrix. The numerical methods and the convergence results that we give for the form (1.1) also apply to (1.5).

The question that arises immediately is: when does there exist a solution to problem (1.1) or (1.2), and when is it unique? The answer has been given for various special cases. We refer the reader to Borg (1946), Hadeler (1968), Hald (1972) and Hochstadt (1974) for some basic results and more references. In the case of the discrete problem (1.3) one can show that if p is symmetric ($p_i = p_{n+1-i}$) then the set $\{\lambda_i^*\}_1^n$ determines $\{p_i\}_1^n$. If p is not symmetric one must provide two sets of eigenvalues corresponding to two sets of boundary conditions.

In this paper we will assume that (1.1) has a solution and will concentrate on how to compute it numerically.

2. Formulation of the Numerical Methods

We will consider only the problem (1.1). We note that stable algorithms exist for some special cases of (1.1); see de Boor and Golub (1978), which treats the construction of a Jacobi matrix from spectral data.

Let $\lambda_1(c) \leq \ldots \leq \lambda_n(c)$ be the eigenvalues of $A(c) = \sum_{i=1}^{n} c_i A_i$. We will consider two formulations of the problem.

Formulation I. Solve the nonlinear system.

$$f(c) = \begin{bmatrix} \lambda_1(c) - \lambda_1^* \\ \\ \lambda_n(c) - \lambda_n^* \end{bmatrix} = 0 \tag{2.5}$$

Formulation II. Solve the nonlinear system

$$g(c) = \begin{bmatrix} \det(A(c) - \lambda_1^* I) \\ \vdots \\ \det(A(c) - \lambda_n^* I) \end{bmatrix} = 0 \tag{2.6}$$

The first formulation is the most natural one and has been used by Downing and Householder (1956) in the case of the additive inverse eigenvalue problem, and by Kublanovskaja (1970) for the general linear form (1.1) The second formulation has been proposed by Biegler-König (1981). We will give some arguments that indicate that the latter has some major disadvantages.

Consider for example the 2×2 case where the $\{A_i\}$ are diagonal. Then

$$A(c) = c_1 \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} + c_2 \begin{bmatrix} b_1 & 0 \\ 0 & b_2 \end{bmatrix}$$

and $\lambda_1(c) = c_1 a_1 + c_2 b_1$, $\lambda_2(c) = c_1 a_2 + c_2 b_2$. Formulation I gives rise to the linear system

$$\begin{aligned} c_1 a_1 + c_2 b_1 - \lambda_1^* &= 0 \\ c_1 a_2 + c_2 b_2 - \lambda_2^* &= 0 . \end{aligned} \tag{2.7}$$

On the other hand, using Formulation II we obtain two quadratic equation

$$\begin{aligned} (c_1 a_1 + c_2 b_1 - \lambda_1^*)(c_1 a_2 + c_2 b_2 - \lambda_1^*) &= 0 \\ (c_1 a_1 + c_2 b_1 - \lambda_2^*)(c_1 a_2 + c_2 b_2 - \lambda_2^*) &= 0 . \end{aligned} \tag{2.8}$$

If we applied Newton's method we would find the solution of (2.7) in

one step, but it would take several steps to solve (2.8) accurately.

The behaviour observed in this example occurs in general. Formulation II complicates the problem and frequently gives rise to badly conditioned systems. The problem is intrinsically simple along some directions. Let

$$\hat{A}(\alpha) = A(\alpha c) = \sum_{i=1}^{n} \alpha c_i A_i = \alpha A(c) .$$

The eigenvalues of $\hat{A}(\alpha)$ are linear functions of α and the eigenvectors are constant. Therefore along any 1-dimensional subspace the problem is linear. Formulation I preserves this property, which helps to accelerate the convergence of the numerical methods. On the other hand it is easy to see that using Formulation II, g is not linear along any 1-dimensional subspace, but its components are in general polynomials of order n.

The first mehtod we will consider consists of applying Newton's method to solve the nonlinear system given by Formulation I. If all the eigenvalues of A(c) are distinct then they are differentiable (see Theorem 1 below) and

$$\frac{\partial \lambda_i(c)}{\partial c_j} = q_i(c)^T A_j q_i(c) , \qquad (2.9)$$

where $q_i(c)$ is an eigenvector of A(c) corresponding to λ_i and $\|q_i(c)\|_2 = 1$. As A(c) is symmetric, $\{q_i\}_1^n$ is an orthonormal set. We can now apply Newton's method to solve (2.5):

Method I

Choose a value for c^0. For k = 0,1,2,...

1. Construct $A(c^k) = \sum_{i=1}^{n} c_i^k A_i$
2. Find the eigenvalues and eigenvectors of A(c):

   $$Q^T A(c) Q = \Lambda$$

   where $Q = [q_1(c^k),\ldots,q_n(c^k)]$ and $\Lambda = \mathrm{diag}(\lambda_i(c^k))$. Let $f_i(c^k) = \lambda_i(c^k) - \lambda_i^*$, i = 1,...,n. Stop if $\|f(c^k)\|_2$ is sufficiently small.
3. Form $J(c^k)$ by

   $$J_{ij} = q_i(c^k)^T A_j q_i(c^k) . \qquad (2.10)$$

4. Compute c^{k+1} by solving

   $$J(c^k)(c^{k+1} - c^k) = -f(c^k) . \qquad (2.11)$$

In the case of the additive inverse eigenvalue problem this method reduces to that of Downing and Householder (1956). Even though it is stated in a different form, it can be seen that the method used for shell model computations in nuclear spectroscopy (see Brussard and Glaudemans (1977)) is also equivalent to this method. Kublanovskaja (1970) gives conditions that guarantee the convergence of Method I, using a line search at step 4.

Now let us consider applying Newton's method to Formulation II. Note that the i-th equation in (2.6) can be written as

$$g_i(c) = \prod_{k=1}^{n} (\lambda_k(c) - \lambda_i^*) .$$

Therefore, using (2.9)

$$\frac{\partial g_i(c)}{\partial c_j} = \sum_{k=1}^{n} \frac{\partial \lambda_k(c)}{\partial c_j} \prod_{\ell \neq k} (\lambda_\ell(c) - \lambda_i^*) = \sum_{k=1}^{n} q_k^T A_j q_k \prod_{\ell \neq k} (\lambda_\ell - \lambda_i^*) . \tag{2.12}$$

Let $G(c)$ be defined by $G_{ij} = \left(\frac{\partial g_i(c)}{\partial c_j}\right)$. Then (2.12) can be written as

$$G(c) = \begin{bmatrix} \prod_{\ell \neq 1} (\lambda_\ell - \lambda_1^*) & \cdots & \prod_{\ell \neq n} (\lambda_\ell - \lambda_1^*) \\ \vdots & & \vdots \\ \prod_{\ell \neq 1} (\lambda_\ell - \lambda_n^*) & \cdots & \prod_{\ell \neq n} (\lambda_\ell - \lambda_n^*) \end{bmatrix} \begin{bmatrix} q_1^T A_1 q_1 & \cdots & q_1^T A_n q_1 \\ \vdots & & \vdots \\ q_n^T A_1 q_n & \cdots & q_n^T A_n q_n \end{bmatrix}$$

$$= \operatorname{diag}(g_i) \begin{bmatrix} \frac{1}{\lambda_1 - \lambda_1^*} & \cdots & \frac{1}{\lambda_n - \lambda_1^*} \\ \vdots & & \vdots \\ \frac{1}{\lambda_1 - \lambda_n^*} & \cdots & \frac{1}{\lambda_n - \lambda_n^*} \end{bmatrix} \cdot J$$

$$= \operatorname{diag}(g_i) \operatorname{diag}\left(\frac{1}{f_i}\right) W J , \tag{2.13}$$

where f_i is the i-th component of (2.5), J is given by (2.10) and the matrix W is defined by

$$W = \begin{bmatrix} 1 & \cdots & \frac{\lambda_1 - \lambda_1^*}{\lambda_n - \lambda_1^*} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ \frac{\lambda_n - \lambda_n^*}{\lambda_1 - \lambda_n^*} & \cdots & 1 \end{bmatrix}$$

Then the Newton step for Formulation II is:

$$\begin{aligned} c^{k+1} &= c^k - G(c^k)^{-1} g(c^k) \\ &= c^k - J^{-1}(c^k) W^{-1}(c^k) \operatorname{diag}(f_i(c^k)) \operatorname{diag}\left(\frac{1}{g_i(c^k)}\right) g(c^k) \\ &= c^k - J^{-1}(c^k) W^{-1}(c^k) f(c^k) . \end{aligned} \tag{2.14}$$

Method II differs then only from Method I in step 4, where $J(c^k)$ should be replaced by $W(c^k)J(c^k)$. This method has been used by Biegler-König (1981) and generalizes an algorithm of Lancaster (1964-a).

Note that if the $\{\lambda_i^*\}_1^n$ are all distinct then $W \to I$ as $\|f(c^k)\|_2 \to 0$. Asymptotically (2.11) and (2.14) coincide in this case, and both methods are usually quadratically convergent. Numerical tests show that Method I converges faster (see the example 1 in Section 5) and that Method II frequently suffers from ill-conditioning. This can be explained by observing that W becomes nearly singular in various situations, particularly far from the solution.

If $\{\lambda_i^*\}_1^n$ contains multiple eigenvalues G becomes singular as we approach the solution. This destroys quadratic convergence. By contrast we show in Section 4 below that Method I is usually quadratically convergent, even in the multiple eigenvalue case.

When the matrices $\{A_i\}$ are not sparse the cost of forming J (see (2.10)) can be very high. One could avoid this problem by using Broyden's method (see Broyden (1965)). However, numerical experiments indicate that Broyden's method generally needs a large number of iterations to approach the solution.

3. Quadratic convergence of Method I when $A(c^*)$ has distinct eigenvalues.

When $\{\lambda_i^*\}_1^n$ are distinct Method I has local quadratic convergence under mild contitions. This follows from standard properties of Newton's method, using the well-known fact that the eigenvalues are differentiable when they are distinct.

Theorem 1. Assume that (1.1) has a solution c^*, i.e. $f(c^*) = 0$, and that $A(c^*)$ has distinct eigenvalues. Then

(a) f is (Fréchet) differentiable in a neighborhood of c^*.

(b) If in addition the Jacobian, which we will henceforth denote by Df, is nonsingular at c^*, then there exists a neighborhood of the solution, $N(c^*)$, such that for all $c^0 \in N(c^*)$ the iterates of Method I converge to c^* and

$$\limsup_{k \to \infty} \frac{\|c^* - c^{k+1}\|}{\|c^* - c^k\|^2} < \infty$$

In other words, Method I has, locally, at least Q-quadratic convergence in the sense of Ortega and Rheinboldt (1970).

Proof: The fact that the eigenvalues of $A(c^*)$ are distinct implies that the eigenvectors are continuous at c^* (see Ortega (1972), p. 54). The partial derivatives of f are given by (2.9) (see Ortega (1972) or Kato (1966), p.81), and are therefore continuous at c^*. It follows that f is differentiable at c^*. It will be shown later that the partial second derivatives of f are given by (4.4), and again, these are continuous, so f is twice differentiable at c^*. The quadratic convergence follows by the well-known properties of Newton's method (see Ortega and Rheinboldt (1970), p.312). □

4. Quadratic Convergence of Method I when $A(c^*)$ has multiple eigenvalues.

Suppose that $A(c^*)$ has multiple eigenvalues. Then the eigenvectors $\{q_i(c^*)\}$ are not unique, and they cannot in general be defined to be continuous functions of c at c^*. Furthermore, the eigenvalues and hence f(c) are not in general differentiable at c^*. These remarks may be verified by considering the example

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad c^* = \lambda^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix} .$$

Note that if c and $\bar{c}$ are such that $A(c)$ and $A(\bar{c})$ have distinct eigenvalues and $\|c - c^*\|$, $\|\bar{c} - c^*\|$ and hence $\|c - \bar{c}\|$ are arbitrarily small, $|q_i(c) - q_i(\bar{c})|$ can be large. It is thus natural to expect that Method I has, at best, slow convergence to c^*, since standard proofs of quadratic convergence require f to be differentiable and Df to be Lipschitz continuous at the solution. It is however a remarkable fact that Method I is usually quadratically convergent to a solution c^* even when $A(c^*)$ has multiple eigenvalues. The reason is essentially that the eigenvalues can be considered as twice differentiable functions of a single variable along any line passing through the solution c^*. Consequently every Newton step produces an excellent estimate of the solution, and the iterates $\{c^*\}$ converge quadratically although $\{Df(c^k)\}$ does not converge.

To prove this we first need the following result found in Rellich (1953).

Lemma 1 (Rellich). Let $B(\alpha)$ be a real symmetric $n \times n$ matrix function of a single variable α. Assume that $B(\alpha)$ is analytic[+] in a neighborhood of $\alpha = 0$, i.e. it can be expanded in a power series convergent in that neighborhood:

$$B(\alpha) = B_0 + \alpha B_1 + \alpha^2 B_2 + \ldots .$$

Let σ be an eigenvalue of $B(0)$ with multiplicity m. Then there exist scalar functions $\sigma_i(\alpha)$ and vector functions $v_i(\alpha)$, $i = 1,\ldots,m$, all analytic in a neighborhood of $\alpha = 0$, such that $\sigma_i(\alpha)$ and $v_i(\alpha)$ are an eigenvalue-eigenvector pair of $B(\alpha)$, i.e.

$$B(\alpha)v_i(\alpha) = \sigma_i(\alpha)v_i(\alpha) \qquad i = 1,\ldots,m .$$

Furthermore $\{v_i(\alpha)\}$ are orthonormal and $\sigma_i(0) = \sigma$, $i = 1,\ldots,m$.

Notice that Lemma 1 does not state that $\sigma_1(\alpha) \leq \ldots \leq \sigma_m(\alpha)$. For example if $B(\alpha) = \begin{bmatrix} \alpha & 0 \\ 0 & -\alpha \end{bmatrix}$ we can set $\sigma_1(\alpha) = \alpha$ and $\sigma_2(\alpha) = -\alpha$. If we wanted $\sigma_1(\alpha) \leq \sigma_2(\alpha)$ we would need to define $\sigma_1(\alpha) = -|\alpha|$, $\sigma_2(\alpha) = |\alpha|$, but these are not analytic. However it is clear that in a one-sided neighborhood of $\alpha = 0$, say $0 \leq \alpha \leq \hat{\alpha}$, we can assume that $\sigma_1(\alpha) \leq \ldots \leq \sigma_m(\alpha)$.

Lemma 1 shows that $\{\sigma_i(\alpha)\}$, a representation of the eigenvalues of $B(\alpha)$, have derivatives of all orders in a neighborhood of $\alpha = 0$. Lemma 2, which is a modification of a result of Lancaster (1964-b), explicitly shows the form of the first and second derivatives.

[+] By a vector or matrix function analytic in α we mean one whose components are analytic in α, i.e. are convergent power series in α.

Lemma 2. Let $B(\alpha)$, $\sigma_i(\alpha)$, $v_i(\alpha)$, $i = 1,\ldots,n$ be as defined in Lemma 1 (applying this lemma once for each distinct eigenvalue of $B(0)$). Let $\Sigma(\alpha) = \text{diag}(\sigma_i(\alpha))$ and $V(\alpha) = [v_1(\alpha),\ldots,v_n(\alpha)]$. Note that $V(\alpha)^T V(\alpha) = I$ and

$$\Sigma(\alpha) = V(\alpha)^T B(\alpha) V(\alpha) \quad . \tag{4.1}$$

Define[†]

$$R(\alpha) = V(\alpha)^T B'(\alpha) V(\alpha) \tag{4.2}$$

with components $\{r_{ij}(\alpha)\}$. Then in a neighborhood of $\alpha = 0$,

$$\sigma_i'(\alpha) = r_{ii}(\alpha) = v_i^T(\alpha)\, B'(\alpha) v_i(\alpha) \tag{4.3}$$

and

$$\sigma_i''(\alpha) = v_i(\alpha)^T B''(\alpha) v_i(\alpha) + 2 \sum_{\substack{k=1 \\ \sigma_k(\alpha) \neq \sigma_i(\alpha)}}^{n} \frac{r_{ik}^2}{\sigma_i(\alpha) - \sigma_k(\alpha)} \tag{4.4}$$

Proof: Differentiating (4.1) and omitting the argument α, we obtain

$$\Sigma' = R + (V^T)'BV + V^T BV' = R + S\Sigma - \Sigma S \tag{4.5}$$

where

$$S = (V^T)'V = -V^T V'$$

with components $\{s_{ij}\}$. Equating the diagonal elements of (4.5) then gives (4.3). Equating the off-diagonal elements of (4.5) we obtain

$$0 = r_{ij}(\alpha) + s_{ij}(\alpha)(\sigma_j(\alpha) - \sigma_i(\alpha)) \quad .$$

Thus we have

$$r_{ij}(\alpha) = 0, \quad i \neq j, \quad \text{if } \sigma_i(\alpha) = \sigma_j(\alpha) \tag{4.6}$$

$$s_{ij}(\alpha) = \frac{r_{ij}}{\sigma_i(\alpha) - \sigma_j(\alpha)}, \quad i \neq j, \quad \text{if } \sigma_i(\alpha) \neq \sigma_j(\alpha) \tag{4.7}$$

Differentiating (4.2) we have

$$R' = V^T B'' V + (V^T)'B'V + V^T B'V' = V^T B'' V + SR - RS \quad .$$

Thus

$$\sigma_i''(\alpha) = r_{ii}'(\alpha) = v_i(\alpha)^T B''(\alpha) v_i(\alpha) + \sum_{k=1}^{n} s_{ik} r_{ki} - r_{ik} s_{ki} \quad .$$

[†] By $B'(\alpha)$ we mean the matrix whose components are obtained by differentiating the components of B with respect to α.

Using (4.6), (4.7) and the symmetry of R we obtain (4.4). □

Note that if B(0) has multiple eigenvalues, say $\sigma_i(0) = \sigma_j(0)$ for some $i \neq j$, but $\sigma_i(\alpha) \neq \sigma_j(\alpha)$ for $\alpha \neq 0$, we must have that $r_{ij} \to 0$ as $\alpha \to 0$, because $r_{ij}(\alpha)$ (and $\sigma_i(\alpha)$) are continuous at $\alpha = 0$ by Lemma 1.

We are now ready to prove the main result.

Theorem 2. Assume that (1.1) has a solution c^*, i.e. $f(c^*) = 0$, and suppose that $A(c^*)$ has multiple eigenvalues. Let c^k be the iterates of Method I. Assume that for all k, $A(c^k)$ has distinct eigenvalues so that $Df(c^k)$ exists, and assume that $Df(c^k)$ is nonsingular so that Method I is well defined. For every vector u on the unit sphere, i.e. $\|u\|_2 = 1$, define

$$B_u(\alpha) = A(c^* + \alpha u)$$

and let $(\sigma_u)_i(\alpha)$, $(v_u)_i(\alpha)$, $i = 1,\ldots,n$ be its eigenvalues and eigenvectors, all analytic in a neighborhood of $\alpha = 0$, as given by Lemma 1. Define $J_u(\alpha)$ by

$$(J_u)_{ij}(\alpha) = (v_u)_i(\alpha)^T A_j (v_u)_i(\alpha) \ . \tag{4.8}$$

Now assume that, for all u with $\|u\|_2 = 1$, $J_u(0)$ is nonsingular and that

$$\sup_{\|u\|_2 = 1} \|J_u(0)^{-1}\| < \infty \ . \tag{4.9}$$

Then there exists a neighborhood of c^*, $N(c^*)$, such that for all $c^0 \in N(c^*)$ the iterates of Method I converge to c^* and

$$\limsup_{k\to\infty} \frac{\|c^* - c^{k+1}\|}{\|c^* - c^k\|^2} < \infty \ , \tag{4.10}$$

i.e. Method I is at least Q-quadratically convergent.

Proof: By the remarks following Lemma 1 we can assume that for all u with $\|u\|_2 = 1$, $(\sigma_u)_1(\alpha) \leq \ldots \leq (\sigma_u)_n(\alpha)$ in a one-sided neighborhood of $\alpha = 0$, say $0 \leq \alpha \leq \hat{\alpha}$, and hence that for such α ,

$$(\sigma_u)_i(\alpha) = \lambda_i(c^* + \alpha u) .$$

Let

$$f_u(\alpha) = f(c^* + \alpha u) = \begin{bmatrix} \lambda_1(c^* + \alpha u) - \lambda_1^* \\ \vdots \\ \lambda_n(c^* + \alpha u) - \lambda_n^* \end{bmatrix} = \begin{bmatrix} (\sigma_u)_1(\alpha) - \lambda_1^* \\ \vdots \\ (\sigma_u)_n(\alpha) - \lambda_n^* \end{bmatrix}$$

Now let c be an iterate c^k generated by Method I; we will omit the superscript for convenience. Define

$$\beta = \|c - c^*\|_2, \quad u = \frac{1}{\beta}(c - c^*) .$$

Note that $\|u\|_2 = 1$, $f(c) = f_u(\beta)$ and that $Df(c)$, which exists by assumption, equals $J_u(\beta)$. Now consider the new iterate $\bar{c}$ given by

$$\bar{c} = c - Df(c)^{-1} f(c) .$$

Using the fact that $f(c^*) = 0$, we have

$$\|\bar{c} - c^*\| \leq \|Df(c)^{-1}\| \, \|Df(c)(c - c^*) - f(c) + f(c^*)\|$$

and hence

$$\|\bar{c} - c^*\| \leq \|J_u(\beta)^{-1}\| \, \|J_u(\beta)\, \beta u - f_u(\beta) + f_u(0)\| . \qquad (4.11)$$

Now by Lemma 2 and the definition (4.8),

$$(f_u)_i'(\alpha) = (\sigma_u)_i'(\alpha) = (v_u)_i(\alpha)^T \Big(\sum_{j=1}^{n} u_j A_j \Big) (v_u)_i(\alpha)$$

$$= (J_u(\alpha)u)_i .$$

From (4.11) we therefore get

$$\|\bar{c} - c^*\| \leq \|J_u(\beta)^{-1}\| \, \|\beta f_u'(\beta) - f_u(\beta) + f_u(0)\| .$$

The key point of the argument now appears: although $f(c)$ is not differentiable at $c = c^*$, $f_u(\alpha)$ is differentiable at $\alpha = 0$ and we can apply the mean value theorem (Ortega and Rheinholdt (1970), p. 70) to give

$$\|\bar{c} - c^*\| \leq \|J_u(\beta)^{-1}\| \sup_{0 \leq \gamma \leq 1} \|f_u'(\beta) - f_u'(\gamma\beta)\| \, \beta .$$

Since $(f_u)_i''(0) = (\sigma_u)_i''$ is defined and given by (4.4) in Lemma 2, we have, if c is close enough to c^*,

$$\|\bar{c} - c^*\| \leq 2\|J_u(\beta)^{-1}\| \, \|f_u''(0)\|\beta^2 \tag{4.12}$$

It is clear from (4.4) that

$$\sup_{\|u\|_2 = 1} \|f_u''(0)\| < \infty.$$

Finally using (4.9), (4.12) and the continuity of $J_u(\alpha)$ we obtain (4.10). This completes the proof of the theorem. □

Although we have given Theorem 2 in terms of the inverse eigenvalue problem, it is clear that it could be stated more generally. The quadratic convergence property for Newton's method thus holds for any function which is not differentiable at the solution, but which does have the appropriate Lipschitz continuity of the derivative along rays with end points at the solution. A different approach to proving Theorem 1 would be to generalize a theorem of Stepleman (1969). His result concerns the convergence of Newton's method for functions which are not differentiable at the solution but whose domain may be divided into a finite number of regions on each of which it is differentiable. For example, both his result and ours show that Newton's method has quadratic convergence when applied to the function $\phi(\alpha) = |\sin\alpha|$, in a neighborhood of $\alpha = 0$.

5. Numerical Examples

We now present two examples to show the numerical behaviour of Method I. The first example has distinct eigenvalues and the second has multiple eigenvalues at the solution. We used a VAX 11/780 at the Courant Mathematics and Computing Laboratory, with approximately 16 decimal digits of accuracy (double precision arithmetic).

Example 1. (Distinct Eigenvalues) Consider problem (1.1) with $n = 4$, $c^* = [1,1,1,1]^T$,

$\lambda_1^* = .3593534136805996$ E+01 $\quad\quad$ $\lambda_2^* = .7291977090804996$ E+01

$\lambda_3^* = .1294937131581000$ E+02 $\quad\quad$ $\lambda_4^* = .4316511745658000$ E+02

$$A_1 = \begin{bmatrix} 5 & 4 & 1 & 1 \\ 4 & 5 & 1 & 1 \\ 1 & 1 & 4 & 2 \\ 1 & 1 & 2 & 4 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 5 & 7 & 6 & 5 \\ 7 & 10 & 8 & 7 \\ 6 & 8 & 10 & 9 \\ 5 & 7 & 9 & 10 \end{bmatrix},$$

$A_3 = I$ and $A_4 = \text{diag}(i)$. The starting point was $c^0 = [.85,.90,1.5,1.3]$. The results obtained using Method I and Method II are as follows:

Iteration	$\|f\|_2$, Method I	$\|f\|_2$, Method II
1	.337 E+01	.337 E+01
2	.216 E-01	.412 E+00
3	.204 E-04	.192 E-01
4	.949 E-11	.615 E-04
5	---	.708 E-09

Both methods exhibit quadratic convergence but Method I is faster. We observed in almost all our tests that Method I took less iterations and had a faster speed of convergence than Method II.

Example 2 (Multiple eigenvalues). Consider problem (1.1) with $n = 6$, $c^* = [1,\ldots,1]$, $\lambda_1^* = \lambda_2^* = \lambda_3^* = 1$,

$\lambda_4^* = .6295300089280376$ E+01 $\qquad \lambda_5^* = 7$

$\lambda_6^* = .2970469991071963$ E+02 . The matrices $A_1,\ldots,A_6$ are defined as follows:

$$A_k = b_k e_k^T + e_k b_k^T \qquad k = 1,\ldots,6,$$

where e_k is the k-th column of the identity matrix and

$$b_1 = \begin{bmatrix} 6 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad b_2 = \begin{bmatrix} 6 \\ 3.5 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad b_3 = \begin{bmatrix} 5 \\ 2 \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad b_4 = \begin{bmatrix} -3 \\ 0 \\ -1 \\ 2 \\ 0 \\ 0 \end{bmatrix}, \quad b_5 = \begin{bmatrix} -4 \\ -4 \\ 0 \\ 2 \\ 3.5 \\ 0 \end{bmatrix}, \quad b_6 = \begin{bmatrix} -9 \\ -4 \\ -3 \\ 5 \\ 6 \\ 6 \end{bmatrix}$$

The starting point was $c^0 = [.9,.9,.9,1.1,1.1,1.1]$. Method I exhibits quadratic convergence:

Iteration	$\|f\|_2$, Method I
1	.146 E+01
2	.624 E+00
3	.492 E+00
4	.207 E+00
5	.498 E+00
6	.984 E-01
7	.384 E+00
8	.420 E-01
9	.250 E-03
10	.278 E-06
11	.794 E-13

Acknowledgements

We would like to thank Percy Deift, Alejandro Frank, Gene Golub, and Olof Widlund for several helpful conversations and suggestions.

References

F.W. Biegler-König (1981), A Newton Iteration Process for Inverse Eigenvalue Problems, Numer. Math. 37, pp. 349-354.

G. Borg (1946), Eine Umkehrung der Sturm-Liouvilleschen Eigenwertaufgabe Acta Math. 78, pp. 1-96.

C.G. Broyden (1965), A class of methods for solving nonlinear simultaneous equations, Math. Comp. 19, pp. 577-593.

P.J. Brussard and Glaudemans P.W. (1977), Shell model applications in nuclear spectroscopy, Elsevier.

C. de Boor and G.H. Golub (1978), The numerically stable reconstruction of a Jacobi matrix from spectral data, Linear Algebra and its Appl. 21, pp. 245-260.

A.C. Downing and A.S. Householder (1956), Some Inverse Characteristic Value Problems, J. Assoc. Comput. Mach. 3, pp. 203-207.

K.P. Hadeler (1968), Ein Inverses Eigenwertproblem, Linear Algebra and its Appl. 1, pp. 83-101.

O. Hald (1972), On discrete and numerical inverse Sturm-Liouville problems, Ph.D. dissertation, Dept. of Math., New York University.

H. Hochstadt (1974), On the construction of a Jacobi matrix from spectral data, Linear Algebra and its Appl. 8, pp. 435-446.

T. Kato (1966), Perturbation theory for linear operators, Springer-Verlag.

W. N. Kublanovskaja (1970), On an approach to the solution of the inverse eigenvalue problem (in Russian), Zapiski naučnych Seminarov Leningradskogo Otdelenija Matematičeskogo Instituta, in V.A. Steklova Akademii Nauk SSSR, pp. 138-149.

P. Lancaster (1964-a), Algorithms for Lambda-Matrices, Numer. Math. 6, pp. 388-394.

P. Lancaster (1964-b), On eigenvalues of matrices dependent on a parameter, Numer. Math. 6, pp. 377-387.

J. M. Ortega (1972), Numerical Analysis, Academic Press.

J. M. Ortega and W.C. Rheinboldt (1970), Iterative solution of nonlinear equations in several variables, Academic Press.

F. Rellich (1953), Perturbations theory of eigenvalue problems, Lecture Notes, Math. Department, New York University.

R. S. Stepleman (1969), Finite dimensional analogues of variational and quasi-linear elliptic Dirichlet problems, Computer Science Tech. Rep. 69-88, University of Maryland.